

Mikrozensus Scientific Use File 1973: Dokumentation und Datenaufbereitung

Lechert, Yvonne; Lüttinger, Paul

Veröffentlichungsversion / Published Version
Verzeichnis, Liste, Dokumentation / list

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Lechert, Y., & Lüttinger, P. (2006). *Mikrozensus Scientific Use File 1973: Dokumentation und Datenaufbereitung*. (ZUMA-Methodenbericht, 2006/01). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-262467>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

ZUMA-Methodenbericht 2006/01

Mikrozensus Scientific Use File 1973
Dokumentation und Datenaufbereitung

Yvonne Lechert / Paul Lüttinger

Januar 2006

ISSN 1610-9953

ZUMA
German Microdata Lab (GML)
Quadrat B2,1
Postfach 12 21 55
68072 Mannheim

Telefon: (0621) 1246 - 253

Telefax: (0621) 1246 - 100

E-mail: mikrodaten@zuma-mannheim.de

INHALTSVERZEICHNIS

1. Einleitung	1
2. Mikrozensus 1973	1
2.1 Kurzbeschreibung	1
2.2 Datenerhebung	2
2.3 Die Stichprobe	2
2.4 Das Erhebungsprogramm	2
3. Der faktisch anonymisierte Mikrozensus 1973 – Das Scientific Use File	3
4. Aufbereitung des Rohdatenfiles zum Mikrozensus Scientific Use File 1973 und Überprüfung der Plausibilität des erstellten Systemfiles	6
4.1 Einlesen des Rohdatenmaterials und Erstellung des Systemfiles	6
4.2 Gewichtung und Hochrechnung zum Vergleich mit den Veröffentlichungen des Statistischen Bundesamtes	7
4.3 Bevölkerungsbegriffe und Abgrenzungen im Mikrozensus	8
4.4 Plausibilitätsprüfung des aufbereiteten Mikrozensus Scientific Use Files 1973	9
Anhang A: Randverteilung ausgewählter Variablen des Scientific Use Files 1973	11
Anhang B: SPSS-Syntax für die Plausibilitätsprüfungen des Scientific Use Files 1973	13
Anhang C: Weiterführende Literatur	15

1. Einleitung

In dem vorliegenden Bericht werden nach einer kurzen Beschreibung des Erhebungsprogramms, der Stichprobe und der Besonderheiten des Mikrozensus 1973 die einzelnen Schritte der Aufbereitung der faktisch anonymisierten Mikrozensusdaten (des Scientific Use Files) dargestellt.

Die Bereitstellung der faktisch anonymisierten Daten als Rohdatenfile in ASCII-Format erfolgt durch das Forschungsdatenzentrum (FDZ) des Landesamtes für Datenverarbeitung und Statistik Nordrhein-Westfalen (Düsseldorf). In Kooperation mit dem FDZ stellt das German Microdata Lab dem Nutzer ausführliche Informationen über das Scientific Use File zur Verfügung und bietet Setups zur Aufarbeitung des Datenmaterials und zur Erstellung von System-Files in Datenformaten gängiger Softwarepakete an. Ausführlichere Informationen über die verfügbaren Mikrozensus Scientific Use Files (vorhandene Variablen, verwendete Systematiken, Fragebogen, Schlüsselverzeichnis, intertemporale Vergleichbarkeit, Mikrodaten Tools etc.) werden darüber hinaus in ZUMA-Methodenberichten zum intertemporalen Vergleich von Mikrozensen und im World Wide Web des German Microdata Lab¹ bereitgestellt und können dort abgerufen werden.

Im Folgenden wird nach einer kurzen Beschreibung des Mikrozensus und des Mikrozensus Scientific Use Files 1973 der Vorgang der Datenaufbereitung beschrieben und die notwendigen Schritte zur Überprüfung der Plausibilität von eingelesenen Daten erläutert.

2. Mikrozensus 1973

2.1 Kurzbeschreibung

Der Mikrozensus ist mit einem Auswahlsatz von 1 Prozent der Bevölkerung die größte jährliche Haushaltsbefragung in Europa. Sie wird seit 1957 in Westdeutschland und seit 1991 in den neuen Bundesländern durchgeführt. Grundlage für die ausführlichen Angaben zu bevölkerungs- und arbeitsmarktstatistischen Strukturdaten des Mikrozensus ist die gesamte wohnberechtigte Bevölkerung, zu der im Mai 1973 rund 62 Millionen Menschen gehörten. Der Mikrozensus ist durch eine hohe Kontinuität im Frageprogramm und aufgrund der Auskunftspflicht durch einen niedrigen Unit-Nonresponse gekennzeichnet. Der Mikrozensus er-

¹ Vgl.: <http://www.gesis.org/Dauerbeobachtung/GML/Service/index.htm>

möglicht differenzierte Analysen auf Personen-, Familien- und Haushaltsebene sowohl im Quer- als auch im Längsschnitt und eignet sich für die Beobachtung von Strukturveränderungen sowie zu tiefgegliederten Analysen.

2.2 Datenerhebung

Die Angaben im Mikrozensus 1973 beziehen sich auf die Berichtswoche vom 6. bis 12. Mai 1973. Die Daten des Mikrozensus 1973 wurden durch die Statistischen Landesämter erhoben. Die Befragungen fanden i.d.R. mündlich statt, ein kleiner Teil erfolgte mittels schriftlicher Auskunft der Befragten. Für alle Beteiligten bestand gemäß der gesetzlichen Regelungen Auskunftspflicht. Nur wenige Fragen waren freiwillig zu beantworten. Der Mikrozensus ist eine Haushaltsstichprobe, d.h. alle Personen in einem Haushalt wurden entweder direkt befragt, oder eine erwachsene Person gab Auskunft über die weiteren (alle minderjährigen) Haushaltsmitglieder. Der Mikrozensus 1973 umfasste auch ausländische Haushalte. Zur Erhebung ihrer Angaben wurde der gleiche Fragebogen verwendet wie bei deutschen Haushalten. Die Daten wurden von den Statistischen Landesämtern aufbereitet und beim Statistischen Bundesamt zu einem Datensatz zusammengestellt.

2.3 Die Stichprobe

Für den Mikrozensus 1973 galt der Stichprobenplan von 1972 (Flächenstichprobe auf der Basis der Volkszählung von 1970). Das stichprobenmethodische Grundkonzept des Mikrozensus 1973 ist eine einstufige Klumpenstichprobe. Die bebaute Fläche der Bundesrepublik wurde auf der Grundlage der Wohnungszählung von 1970, die im Rahmen der Volkszählung stattfand, in etwa gleich große Flächen (Auswahlbezirke bzw. Segmente) aufgeteilt. Die Neubautätigkeit wurde dabei berücksichtigt. Ein Prozent dieser Auswahlbezirke gelangen in die Stichprobe. Innerhalb eines Auswahlbezirkes wurden alle Haushalte befragt. Grundlage für die Auswahlbezirke sind ganze Gebäude oder bei größeren Gebäuden Gebäudeteile. Die Gebäude der Auswahlbezirke wurden dabei nach der Zahl ihrer Wohnung in drei Größenklassen bzw. Schichten eingeteilt, zusätzlich wurden Gebäude mit vermuteten Gemeinschaftsunterkünften einer Sonderschicht zugeordnet. Ein Auswahlbezirk umfasste durchschnittlich 23 Wohnungen.

2.4 Das Erhebungsprogramm

Das Fragenprogramm des Mikrozensus besteht aus einem gleich bleibenden Grundprogramm, das sozioökonomische Grundinformationen liefert und die laufende Beobachtung des Ar-

beitsmarktes ermöglicht. Im Grundprogramm, das sich an alle Haushalte richtet, sind 1973 folgende Bereiche enthalten:

- Fragen zum Haushalt/Wohnung
- Fragen zu den Personen im Haushalt
- Staatsangehörigkeit
- Vertriebenen/Flüchtlingsstatus
- Schulbesuch
- Erwerbsbeteiligung
- Jahr der Eheschließung
- Zahl und Alter der Kinder im HH und Familie
- Erste gegenwärtige Erwerbstätigkeit
- Zweite Erwerbstätigkeit
- Arbeitsuche/Arbeitsplatzwechsel
- Unterhalt, Einkommen
- Altersvorsorge
- Krankenversicherung

Als Zusatzprogramm wurden 1973 folgende Merkmale erfasst:

mit einem Auswahlatz von 1 Prozent:

- Urlaubs- und Erholungsreisen
- Vermögenswirksame Leistungen
- Behinderungen von Kindern
- Krankheiten und Unfälle

mit einem Auswahlatz von 0,1 Prozent:

- Arbeitszeit am Wochenende
- Nacht- und Sonntagsarbeit
- Ausbildungsabsichten der Eltern für ihre Kinder
- Krankheiten und Unfälle
- berufliche Mobilität der Frauen
- Einkaufsmöglichkeiten und Einkaufsverhalten in Großstädten
- Ziel und Dauer von Kurzreisen

Im Scientific Use File des MZ73 sind die Fragen des Zusatzprogramms nicht enthalten.

3. Der faktisch anonymisierte Mikrozensus 1973 – Das Scientific Use File

Bei den vorliegenden Daten des Mikrozensus Scientific Use File 1973 handelt es sich um eine faktisch anonymisierte 70-Prozent-Substichprobe des Mikrozensus 1973. Für die faktisch anonymisierte Substichprobe gilt, dass sie als systematische Zufallsauswahl aus dem Originalmaterial des Statistischen Bundesamtes gezogen wurde. Das Datenmaterial wurde dazu

nach Bundesland, Regierungsbezirk, Gemeindegrößenklasse, Anzahl der Personen in Privathaushalten, Auswahlbezirksnummer und laufenden Nummer des Haushaltes im Auswahlbezirk sortiert. Durch diese Anordnung ist gewährleistet, dass die Substichprobe hinsichtlich dieser Merkmale nur geringe zufallsbedingte Abweichungen aufweist. Die Haushalte werden bundesweit aufsteigend neu durchnummeriert, dabei werden Anstaltspersonen wie Einpersonenhaushalte behandelt. Zur Stichprobenziehung diente die letzte (Einer-)Stelle der Haushaltsnummer. Alle Haushalte, deren Einerstelle einer von drei zufällig gezogenen einstelligen Zahl entsprach (3,5,9), gelangten nicht in die Stichprobe.

Weiterhin unterscheidet sich das Scientific Use File vom Originalfile des Mikrozensus dadurch, dass bestimmte Variablen im Scientific Use File, bedingt durch die Anonymisierung, in klassierter und vergrößerter Form verfügbar sind.

So sind im Scientific Use File die Regionalangaben Bundesland und eine vergrößerte Gemeindegrößenklasse (EF6) enthalten. Die folgende Abbildung zeigt die Variablen Bundesland und Gemeindegrößenklasse in der Gliederung, wie sie im Originaldatensatz enthalten ist.

EF1 Land der Bundesrepublik		EF6 Gemeindegrößenklasse (Original)	
Schleswig-Holstein	1	unter 2 000 Einwohner	1
Hamburg	2	2 000 bis unter 5 000 Einwohner	2
Niedersachsen	3	5 000 bis unter 10 000 Einwohner	3
Bremen	4	10 000 bis unter 20 000 Einwohner	4
Nordrhein-Westfalen	5	20 000 bis unter 50 000 Einwohner	5
Hessen	6	50 000 bis unter 100 000 Einwohner	6
Rheinland-Pfalz	7	100 000 bis unter 200 000 Einwohner	7
Baden-Württemberg	8	200 000 bis unter 500 000 Einwohner	8
Bayern	9	500 000 und mehr	9
Saarland	10	unbekannt	@
Berlin (West)	11		

Aus Anonymisierungsgründen muss bei jedem Bundesland jede Größenklasse mindestens 400.000 Einwohner (auf die Gesamtbevölkerung hochgerechnet) beinhalten. Folglich waren entsprechende Zusammenfassungen notwendig, die durch eine Kombination der Variablen Bundesland (EF1) und Gemeindegrößenklasse (EF6) erreicht wurde. Die neu gebildete Variable EF6 ist der folgenden Übersicht zu entnehmen. Die Zusammenfassungen zur Anonymisierung sind in den Kategorien 10-20 der Variablen EF6 beschrieben. Tabelle 1 zeigt die Verteilung der Befragten in den einzelnen Bundesländern auf die verschiedenen Gemeindegrößenklassen der neu gebildeten Variablen EF6.

EF6: Gemeindegrößenklasse (im Scientific Use File 1973)

1	unter 2 000 Einwohner	
2	2 000 bis unter 5 000 Einwohner	
3	5 000 bis unter 10 000 Einwohner	
4	10 000 bis unter 20 000 Einwohner	
5	20 000 bis unter 50 000 Einwohner	
6	50 000 bis unter 100 000 Einwohner	
7	100 000 bis unter 200 000 Einwohner	
8	200 000 bis unter 500 000 Einwohner	
9	500 000 und mehr Einwohner	
10	ohne Aufteilung	$EF1 = 04$
11	unter 2000 Einwohnern oder unbekannt	$EF1 = 01 / EF6 = 1, @$
12	unter 5000 Einwohnern	$EF1 = 05,06 / EF6 = 1,2,3$
13	unter 10000 Einwohnern	$EF1 = 10 / EF6 = 1,2,3$
14	2000 bis unter 10000 Einwohnern	$EF1 = 01 / EF6 = 2,3$
15	10000 bis unter 100000 Einwohnern	$EF1 = 07 / EF6 = 4,5,6$
16	10000 bis unter 200000 Einwohnern	$EF1 = 10 / EF6 = 4,5,7$
17	20000 bis unter 100000 Einwohnern	$EF1 = 01 / EF6 = 5,6$
18	50000 bis unter 200000 Einwohnern	$EF1 = 06 / EF6 = 6,7$
19	100000 bis unter 500000 Einwohnern	$EF1 = 08,09 / EF6 = 7,8$
20	200000 und mehr Einwohnern	$EF1 = 03 / EF6 = 8,9$

Tabelle 1: Bundesland x Gemeindegrößenklasse

		GEMEINDEGRÖßENKLASSEN																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Gesam.
B U N D E S L A N D	1	0	0	0	2823	0	0	0	3639	0	0	4602	0	0	3714	0	0	3964	0	0	0	18742
	2	0	0	0	0	0	0	0	0	12563	0	0	0	0	0	0	0	0	0	0	0	12563
	3	9881	6632	6939	6497	8169	3926	5516	0	0	0	0	0	0	0	0	0	0	0	0	5322	52882
	4	0	0	0	0	0	0	0	0	0	5180	0	0	0	0	0	0	0	0	0	0	5180
	5	0	0	8754	14745	26529	17231	12999	18662	19660	0	0	5513	0	0	0	0	0	0	0	0	124093
	6	0	0	6971	7534	4371	0	0	3467	4714	0	0	8723	0	0	0	0	0	4015	0	0	39795
	7	8783	4126	3093	0	0	0	4608	0	0	0	0	0	0	0	6339	0	0	0	0	0	26949
	8	8723	10547	11509	8291	10092	6122	0	0	4270	0	0	0	0	0	0	0	0	0	6985	0	66539
	9	18625	13751	10000	9877	5098	4642	0	0	12969	0	0	0	0	0	0	0	0	0	4004	0	78966
	10	0	0	0	0	0	0	0	0	0	0	0	0	4791	0	0	3221	0	0	0	0	8012
	11	0	0	0	0	0	0	0	0	14645	0	0	0	0	0	0	0	0	0	0	0	14645
Ges		46012	35056	47266	49767	54259	31921	23123	25768	68821	5180	4602	14236	4791	3714	6339	3221	3964	4015	10989	5322	448366

Auch bei den weiteren Variablen des Scientific Use Files werden je nach Bedarf Merkmalsvergrößerungen vorgenommen, so dass jede Ausprägung in den univariaten Verteilungen des Originalmaterials des Mikrozensus 1973 mindestens **5.000** Fälle in der Grundgesamtheit aufweist.

Die Ausprägungen der Variable Nationalität wurden so aggregiert, dass jede Nationalitätengruppe in der Grundgesamtheit des Originalmaterials wenigstens **50.000** Einwohner umfasst.

4. Aufbereitung des Rohdatenfiles zum Mikrozensus Scientific Use File 1973 und Überprüfung der Plausibilität des erstellten Systemfiles

4.1 Einlesen des Rohdatenmaterials und Erstellung des Systemfiles

Die vom German Microdata Lab bereitgestellten Setups für das Mikrozensus Scientific Use File 1973 dienen zum Einlesen des Rohdatenmaterials und zum Erstellen von Systemfiles für die Programmpakete SPSS, SAS und STATA. Sie beinhalten Programmanweisungen zum Umwandeln von nicht-numerischen Variablen in numerische Variablen, zur Definition von fehlenden Werten sowie zum Versehen der Variablen und ihrer Ausprägungen mit entsprechenden Labeln.

Bei der Aufbereitung der Rohdaten wurden fehlende Werte möglichst in 0 umgewandelt. War dieser Wert schon vergeben, dann wurden die fehlenden Werte in 9 oder 8 (bzw. 99, 999, 98) umkodiert.

Um eine reibungsfreie Aufbereitung der Daten zu gewährleisten, empfiehlt es sich, die in den Setups für das jeweilige Programmpaket vorgesehenen Voreinstellungen zu berücksichtigen und je nach Leistung des verarbeitenden Systems die Setups schrittweise durchführen zu lassen. Sind die Daten eingelesen und ist das entsprechende Systemfile erstellt, können die Verteilungen ausgewählter Variablen des Scientific Use Files 1973 (vgl. Anhang A) zur Überprüfung des fehlerfreien Ablaufs der Datenaufbereitung herangezogen werden.

Als Referenz zur Plausibilitätsprüfung der Scientific Use Files dienen die in den Fachserien des Statistischen Bundesamtes veröffentlichten Ergebnisse des jeweiligen Mikrozensus. Für den Mikrozensus 1973 ist dies die Fachserie A: Bevölkerung und Kultur, Reihe 6, Erwerbstätigkeit. Beim Vergleich der Ergebnisse des Scientific Use Files mit den entsprechenden Veröffentlichungen des Statistischen Bundesamtes ist darauf zu achten, dass die in diesen Reihen veröffentlichten Zahlen gewichtete, auf die Gesamtbevölkerung hochgerechnete Ergebnisse

des Mikrozensus sind. Weiterhin beziehen sich die Zahlen auf nach bevölkerungs- oder erwerbstatistischen Konzepten definierte (Teil-) Populationen, die in der jeweiligen Reihe der Fachserien des Statistischen Bundesamtes erläutert werden.

Im Folgenden werden die Modalitäten der Gewichtung und Hochrechnung der Scientific Use Files sowie die für den Vergleich mit den Veröffentlichungen des Statistischen Bundesamtes relevanten Bevölkerungsbegriffe und Abgrenzungen erläutert. Die konkreten Schritte und Ergebnisse der Plausibilitätsprüfung der aufbereiteten Mikrozensus Scientific Use Files 1973 beim German Microdata Lab werden in Abschnitt 4.4 dargestellt.

4.2 Gewichtung und Hochrechnung zum Vergleich mit den Veröffentlichungen des Statistischen Bundesamtes

Im Scientific Use File 1973 liegen keine Gewichte zur Kompensation der systematischen Ausfälle und des zufallsbedingten Stichprobenfehlers vor.

Zur Anpassung der Daten an die Bevölkerungsfortschreibung wurde für den MZ73 das Verfahren des zufälligen Doppelns und Streichens von Personen vorgenommen. Sollen die Ergebnisse von Häufigkeiten des Scientific Use Files mit Veröffentlichungen des Statistischen Bundesamtes zum Mikrozensus 1973 verglichen werden, muss diese Variable der gedoppelten/gestrichenen Fälle (EF76) bei Auswertungen verwendet werden. Dies geschieht bei der Abgrenzung der verschiedenen Bevölkerungskonzepte, wie z.B. Wohnbevölkerung (siehe Abschnitt 4.3).

Zum Verfahren des Doppelns und Streichens: Den Ist-Zahlen der Stichprobe werden die Soll-Zahlen gegenübergestellt, die sich aus der Bevölkerungsfortschreibung des betreffenden Jahres - gegliedert nach den Merkmalen Regierungsbezirk, Deutsche-Ausländer, Geschlecht - ergeben. Durch das (zufällige) Doppelns und Streichen von Personen (bzw. aller Personen eines Haushaltes) des Mikrozensus wird eine Übereinstimmung mit den Soll-Zahlen der Bevölkerungsfortschreibung herbeigeführt. Dadurch wird zwar strenggenommen nur eine Anpassung an die Randverteilungen der verwendeten Variablen erreicht. Man geht dabei jedoch von der Annahme aus, dass auch für andere Untergliederungen Genauigkeitsverbesserungen erreicht werden.²

Um auf ein Prozent der Bevölkerung, also auf den vollen Auswahlsatz des Mikrozensus hochzurechnen, wird das Gewicht mit dem Kehrwert der Ziehungswahrscheinlichkeit für das Scientific Use File (100/70) multipliziert:

² Erläuterungen zu den Fehlerrechnungen zum Mikrozensus finden sich in den einzelnen Fachserien des Statistischen Bundesamtes.

compute gewicht = 100/70.
weight by gewicht.

Wie erwähnt, sind die in den Fachserien des Statistischen Bundesamtes veröffentlichten absoluten Zahlen gewichtete, auf die Gesamtbevölkerung der Bundesrepublik hochgerechnete Daten des Mikrozensus. Zur Herstellung der Vergleichbarkeit zwischen den absoluten Häufigkeiten des Scientific Use Files mit den Veröffentlichungen des Statistischen Bundesamtes müssen somit die Ergebnisse eigener Berechnungen mit dem Scientific Use File noch zusätzlich mit 100 (Kehrwert der Ziehungswahrscheinlichkeit für den vollen Mikrozensus) multipliziert werden.

compute gewicht = 100*100/70.
weight by gewicht.

4.3 Bevölkerungsbegriffe und Abgrenzungen im Mikrozensus

Zum Vergleich der Ergebnisse des Scientific Use Files (SUF) mit den vom Statistischen Bundesamt veröffentlichten Zahlen sind die jeweiligen Bevölkerungsbegriffe und Erwerbskonzepte zu berücksichtigen. Um solche Vergleiche zu erleichtern, gibt die folgende Tabelle einen Überblick über die Bevölkerungsbegriffe und deren Abgrenzung. Die Variable EF76 (Doppelung/Streichung) muss dabei verwendet werden, um eine Anpassung an die Bevölkerungsfortschreibung zu erzielen.

Tabelle 2: Bevölkerungsbegriffe, Abgrenzungen und Gewichtungen im Mikrozensus SUF 1973

Bevölkerungsbegriff Subpopulation	Filter zur Abgrenzung der Subpopulation
Wohnbevölkerung	(EF58=1 or EF58=2 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0)
Bevölkerung in Privathaushalten	(EF58=1 or EF58=2 or EF58=3 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0) and EF60=1
Bevölkerung am Familienwohnsitz	(EF58=1 or EF58=2 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0) and EF61=1
Erwerbstätige	(EF58=1 or EF58=2 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0) and (EF65<=3 or EF65=9 or EF65=10)
Erwerbslose	(EF58=1 or EF58=2 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0) and (EF65=4 or EF65=5 or EF65=6)
Nichterwerbspersonen	(EF58=1 or EF58=2 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0) and (EF65=7 or EF65=8)

4.4 Plausibilitätsprüfung des aufbereiteten Mikrozensus Scientific Use Files 1973

Die Plausibilitätsprüfung des Mikrozensus Scientific Use Files 1973 erfolgte im German Microdata Lab durch den Vergleich des aufbereiteten Datensatzes mit den vom Statistischen Bundesamt veröffentlichten Tabellen in den entsprechenden Reihen der Fachserie A. Dabei wurde nach den vom Statistischen Bundesamt verwendeten Bevölkerungs- und Erwerbskonzepten gewichtet und selektiert. Tabelle 2 enthält die Liste der zur Plausibilitätsprüfung herangezogenen Variablen, Angaben zu notwendigen Gewichtungen und Abgrenzungen sowie die entsprechenden Seitenzahlen der jeweiligen Datenreihe des Statistischen Bundesamtes.

Tabelle 3: Gewichtung, Abgrenzung und Vergleich mit den Fachserien des Statistischen Bundesamtes für ausgewählte Variablen des Mikrozensus Scientific Use File 1973 (siehe Anhang B)

Fachserie (A) / Reihe 6 'Entwicklung der Erwerbstätigkeit
<p>Wohnbevölkerung: (EF58=1 or EF58=2 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0) Wohnbevölkerung und Erwerbspersonen sowie Erwerbsquoten nach Alter und Familienstand</p> <p>Tabelle 9 (S.32) EF20 Alter (gruppiert) EF21 Familienstand</p> <p>Tabelle 1 (S.14) EF65 Erwerbskonzept (recodiert) EF18 Geschlecht EF65 überwiegender Lebensunterhalt</p> <p>Tabelle 21 (S.53) EF65 Erwerbskonzept (recodiert) EF1 Bundesland</p>
<p>Erwerbstätige select if (EF58=1 or EF58=2 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0) and (Erwerb=1). (siehe Tabellenprogramm: Tabelle E 2.1; Vorbedingung) FS (A) Bevölkerung + Kultur / Reihe 6 'Entwicklung der Erwerbstätigkeit'</p> <p>Tabelle 1 (S. 20) ef39 Stellung im Beruf (recodiert) ef44 Tats. gel. Arbeitszeit (recodiert) ef 65 Bevölkerungstyp nach Erwerbs- und Unterhaltskonzept (Teilpopulation der Erwerbstätigen)</p> <p>Tabelle 12 (S. 35) ef55 Nettoeinkommen</p>

Die Programmanweisungen für die durchgeführten Plausibilitätsprüfungen sind als SPSS-Syntax im Anhang B aufgeführt. Bedingt durch die Ziehung der 70%-Stichprobe ergeben sich Abweichungen zwischen den Häufigkeiten der Variablen des Scientific Use Files und den in Fachserien des Statistischen Bundesamtes veröffentlichten Zahlen. Die Mehrzahl der oben aufgeführten Variablen weichen i.d.R. nur in geringem Maße (meist 0 bis 1 Prozent, max. 5 Prozent) von den veröffentlichten Daten ab. Abweichungen über 5 Prozent wurden bei Vari-

ablen festgestellt, deren Merkmalsausprägungen mit sehr kleinen Fallzahlen besetzt sind (z.B. Staatsangehörigkeit und Wirtschaftsbereiche).

Die Überprüfung der Angaben auf Personenebene mit den Angaben in den Bandsatzergänzungen in Bezug auf die Variablen Wirtschaftszweig des Haushaltsvorstands und Wirtschaftszweig der Familienbezugsperson ergab, dass bei insgesamt 7 Fällen die Angaben zum Wirtschaftszweig (ef41) und die Angaben in der Bandsatzergänzung Wirtschaftszweig des Haushaltsvorstandes (ef100) sowie Wirtschaftszweig des Familienvorstandes (ef122) voneinander abweichen.

Anhang A: Randverteilung ausgewählter Variablen des Scientific Use Files 1973

Bevölkerungskonzept	Wohnbevölkerung	Bevölkerung in Privathaushalten	Bevölkerung am Familienwohnsitz	Erwerbstätige Bevölkerung	Fallzahlen
Merkmal	<i>select if</i> (EF58=1 or EF58=2 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0)	<i>select if</i> (EF58=1 or EF58=2 or EF58=3 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0) and EF60=1	<i>select if</i> (ef61=1)	<i>select if</i> (EF58=1 or EF58=2 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0) (EF65<=3 or EF65=9 or EF65=10)	<i>ohne Selektion und ohne Anpassung</i>
Insgesamt	618 944	614 559	606 637	270 407	448 366
EF17: Staatsangehörigkeit					
Deutscher/Deutschland+Ausland	588426	586049	578333	251003	426555
Algerien/Marokko/Portugal/Tunesien	1019	910	909	700	735
Belgien/Dänemark/Frankr./Luxemb.	667	653	643	331	479
Griechenland	3716	3637	3629	2484	2625
GB/Nordirl./Irl./Norw./Polen/Schwed	571	557	551	289	417
Italien	5314	5097	5074	3344	3764
Jugoslawien	4767	4324	4293	3663	3403
Niederlande	974	957	954	460	721
Österreich	1313	1291	1279	744	943
Schweiz/Tschechoslowakei/Ungarn	547	591	527	304	400
Spanien	2291	2000	1993	1527	1638
Türkei	6657	6099	6060	4290	4738
Vereinigte Staaten von Amerika	591	591	579	174	427
Übriges Ausland	1377	1179	1154	670	996
Staatenlos	713	674	660	423	525
EF18: Geschlecht					
Männlich	296254	294696	289774	171013	217000
Weiblich	322690	319863	316863	99394	231366
EF20: Alter (Gruppiert)					
76 und älter	30776	28880	28863	1181	21791
70 - u. 75	26734	26216	26194	2421	18943
65 - u. 70	34383	33993	33941	5777	24487
60 - u. 65	37540	37264	37141	15514	26996
55 - u. 60	25193	24940	24827	14820	18238
50 - u. 55	37393	37164	36983	24966	27227
45 - u. 50	38883	38654	38427	28436	28336
40 - u. 45	38974	38654	38420	29109	28531
35 - u. 40	47201	46766	46461	35321	34441
30 - u. 35	47383	46997	46621	35073	34525
25 - u. 30	36619	36470	35340	26936	27218
20 - u. 25	39986	41461	37853	30277	30535
15 - u. 20	43673	43317	42127	20490	31767
unter 15	134207	133781	133437	86	95331
EF21: Familienstand					
Ledig	241520	239399	233450	67586	176129
Verheiratet	312366	312039	310107	187203	225952
Verwitwet	53439	51900	51881	7867	37911
Geschieden	11620	11221	11199	7751	8374

EF39: Stellung im Beruf					
Selbst. ohne Beschäftigte	13727	13790	13703	13713	12285
Selbst. mit Beschäftigten	11559	11634	11524	11546	8749
Mithelf. Familienang.	14900	14893	14873	14893	11372
Beamter, Richter, Soldat	15556	15869	15483	15549	11510
Angestellter	82579	82489	81074	82221	60252
Arbeiter/Heimarbeiter	116530	114791	113720	115686	83611
Kaufm.-techn. Auszubild.	5146	4974	4809	5136	3753
Gewerbl. Auszubildender	6457	6460	6246	6453	4713
Wehrpflichtiger	2427	2427	2404	2427	1711
Berufssoldat	2784	2260	2243	2784	1980
Entfällt	347280	344971	340559	-	248430
EF48: Überw. Lebensunterhalt					
Erwerbstätigkeit	241837	240301	237163	241837	178509
Arbeitslosengeld/-Hilfe	749	749	746	14	530
Rente, Pension	90350	88094	87890	4009	64065
Unterhalt durch Eltern etc.	267594	269580	265311	18186	191937
Eigenes Vermögen	4711	4653	4637	814	3388
Sozialhilfe	8491	6494	6243	336	6230
Soldat	5211	4687	4647	5211	3707
EF55: Nettoeinkommen Person					
Angabe fehlt	12179	9653	9417	6069	8972
Unter 150 DM	11214	10687	10493	3113	8108
150 bis unter 300 DM	28237	28020	27533	15671	20349
300 bis unter 450 DM	28231	27753	27089	12241	20364
450 bis unter 600 DM	37596	36951	36434	17486	27016
600 bis unter 800 DM	49873	48871	48267	29789	36185
800 bis unter 1000 DM	57350	56657	56094	44463	42115
1000 bis unter 1200 DM	49369	48930	48479	42969	36234
1200 bis unter 1400 DM	30801	30704	30389	27617	22516
1400 bis unter 1600 DM	17633	17683	17483	15971	12934
1600 bis unter 1800 DM	10417	10511	10339	9456	7670
1800 bis unter 2000 DM	7356	7431	7297	6703	5435
2000 bis unter 2200 DM	5103	5156	5071	4700	3792
2200 bis unter 2500 DM	3894	3937	3877	3620	2874
2500 bis unter 3000 DM	3577	3629	3536	3387	2677
3000 und mehr	6009	6113	5959	5641	4576
Mith. Fam./Selbst. Landwirt	21511	21514	21481	21511	15615
Kein Einkommen	238594	240357	237400	-	170934
EF57: Personentyp					
Bis 7 Jahre ohne Schulbesuch	55781	55673	55603	-	39620
Schüler bis 14 Jahre	83969	83631	83317	-	59661
Schüler/Studenten über 14 Jahre	23703	25393	21583	-	18853
Erwerbstätige	270407	268311	264836	270407	199031
Sonstige	185084	181550	181299	-	131201
EF84: Vollzeit-/Teilzeittätigkeit					
Angabe fehlt	348537	346247	341801	-	249335
Vollzeittätigkeit	243839	241824	238433	243839	179949
Teilzeittätigkeit	26569	26487	26403	26569	19082

Anhang B: SPSS-Syntax für die Plausibilitätsprüfungen des Scientific Use Files 1973

- * Plausibilitätsprüfung:
- * Vergleich mit den Fachserien des Statistischen Bundesamtes MZ1973
- * Vorarbeiten: Recodierungen einzelner Variablen

*** Erwerbskonzept

```
Recode EF65 (1 thro 3=1) (9 thro 10=1) (4 thro 6=2) (else= 3) into Erwerb.  
Var Lab Erwerb 'Erwerbsstatus'.  
Val Label Erwerb 1 'Erwerbstätige'  
                2 'Erwerbslose'  
                3 'Nichterwerbspersonen'.
```

*** überwiegender Lebensunterhalt

```
Recode EF65 (1 9 10=1) (2 4 =2) (5 7=3) (3 6 8= 4) into Unterh.  
Var Lab Unterh 'Lebensunterhalt'.  
Val Label Unterh  
1 'Erwerbstätigkeit'  
2 'ALH'  
3 'Rente'  
4 'Angehörige'.
```

*** Tatsächlich geleistete Stunden

```
Recode EF44 (0=0) (1 thro 14=1) (15 thro 23=2) (24 thro 39=3) (40 thro 41= 4)  
(42 thro 44=5) (45=6) (46 thro 47=7) (48=8) (49 thro 54=9)  
(55 thro 69=10) (else=11) into Stunden.  
Var Lab Stunden 'tats. gel. Arbeitsstunden'.  
Val Label Stunden  
0 '0 Stunden'  
1 '1-14 Stunden'  
2 '15-23 Stunden'  
3 '24-39 Stunden'  
4 '40-41 Stunden'  
5 '42-44 Stunden'  
6 '45 Stunden'  
7 '46-47 Stunden'  
8 '48 Stunden'  
9 '49-54 Stunden'  
10 '55-69 Stunden'  
11 '70 unde mehr Stunden'.
```

*** Stellung im Beruf

```
Recode EF39 (0 thro 1=1) (2 thro 2=2) (3 thro 9=3) into Beruf.  
Var Lab Beruf 'Stellung im Beruf'.  
Val Label Beruf 1 'Selbständige'  
2 'Mithelfende'  
3 'Abhängige'.
```

*** Alter (gruppiert) / Tabelle 9 (S.32)

```
Recode EF20 (79 thro 98=1) (0 thro 3=2) (99=2) (4 thro 8=3) (9 thro 13=4) (14 thro 18=5) (19 thro 23=6) (24 thro 28=7)  
(29 thro 33=8) (34 thro 38=9) (39 thro 43=10) (44 thro 48=11) (49 thro 53=12) (54 thro 58=13) (else=14) into Alter.  
Var Lab Alter 'Altersgruppen'.  
Val Label Alter 1 '75 und mehr Jahre'  
2 '70 bis u. 75 Jahre'  
3 '65 bis u. 70 Jahre'  
4 '60 b.u. 65 Jahre'  
5 '55 b.u. 60 Jahre'  
6 '60 b.u. 55 Jahre'  
7 '45 b.u. 50 Jahre'  
8 '40 b.u. 45 Jahre'  
9 '35 b.u. 40 Jahre'  
10 '30 b.u. 35 Jahre'
```


11 '25 b.u. 30 Jahre'
 12 '20 b.u. 25 Jahre'
 13 '15 b.u. 20 Jahre'
 14 'unter 15 Jahren'.

*** **Nettoeinkommen der Person / Tabelle 12 (S.35)**

Recode ef55 (1=1) (2=2) (3 thro 4=3) (5=4) (6 thro 7=5) (8 thro 10=6) (11 thro 15=7) into Einko.
 Var lab Einko 'Nettoeinkommensgruppen'.
 Val lab Einko 1 'unter 150 DM'
 2 '150 b.u. 300 DM'
 3 '300 b.u. 600 DM'
 4 '600 b.u. 800 DM'
 5 '800 b.u. 1200 DM'
 6 '1200 b.u. 1800 DM'
 7 '1800 und mehr'.

*** **Wirtschaftsbereiche (S. 20 / Tabelle 1)**

Recode EF41 (1 thro 5=1) (7 thro 11=2) (12 thro 44=2) (45 thro 47=2) (48 thro 56=3) (57 thro 63=3)
 (64 thro 65=4) (66 thro 85=4) (86 thro 90=4) (91 thro 95=4) (else=5) into Wirt.
 Var Lab Wirt 'Wirtschaftsbereiche'.
 Val Label Wirt 1 'Landwirtschaft etc.'
 2 'Produzierendes Gewerbe'
 3 'Verkehr/Handl.'
 4 'sonstige Dienstl.'
 5 'ohne Angabe'.

*+++++.
 *+++ **FS (A) / Reihe 6 'Entwicklung der Erwerbstätigkeit (S. 32 / Tabelle 9)**
 *+++ **Wohnbevölkerung**
 *+++ Erwerbspersonen sowie Erwerbsquoten nach Alter und Familienstand 1973
 *+++
 *+++ Tabelle 9 (S. 32) ef20 Alter (gruppiert) - Geburtsjahr (generiert)
 *+++ ef21 Familienstand
 *+++ Tabelle 1 (S. 14) ef65 Erwerbskonzept (recodiert)
 *+++ ef18 Geschlecht
 *+++ ef65 überwiegender Lebensunterhalt
 *+++ Tabelle 21 (S. 53) ef1 Bundesland
 *+++ ef65 Erwerbs- und Unterhaltskonzept
 *+++++.

Compute gewicht=100/70.
 weight by gewicht.
 temporary.
 select if (EF58=1 or EF58=2 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0).
 fre Alter ef21 ef18 Erwerb Unterh ef65 ef1.

*+++++.
 *+++ **Erwerbstätige**
 *+++ (siehe Tabellenprogramm: Tabelle E 2.1; Vorbedingung)
 *+++ FS (A) Bevölkerung + Kultur / Reihe 6 'Entwicklung der Erwerbstätigkeit'
 *+++
 *+++ Tabelle 1 (S. 20) ef39 Stellung im Beruf (recodiert)
 *+++ ef44 Tats. gel. Arbeitszeit (recodiert)
 *+++ ef 65 Bevölkerungstyp nach Erwerbs- und Unterhaltskonzept
 *+++ (Teilpopulation der Erwerbstätigen)
 *+++ ef41 Wirtschaftszweig (recodiert)
 *+++ Tabelle 12 (S. 35) ef55 Nettoeinkommen
 *+++++.

Compute gewicht=100/70.
 weight by gewicht.
 temporary.
 select if (EF58=1 or EF58=2 or EF58=4) and (EF76=1 or EF76=3 or EF76=0) and (EF38=1 or EF38=0) and (Erwerb=1).
 fre Stunden Beruf Einko Wirt.

Anhang C: Weiterführende Literatur

Informationen zum Mikrozensus und seiner Konzeption:

- Emmerling, Dieter/Riede, Thomas (1997): 40 Jahre Mikrozensus. *Wirtschaft und Statistik* (3):160-174.
- Hartmann, P. (1989): Der Mikrozensus als Datenquelle für die Sozialwissenschaften. *ZUMA-Nachrichten* Nr. 24, Mai 1989: 6-25.
- Heidenreich, H.-J.(1994): Hochrechnung des Mikrozensus ab 1990. In: Gabler, S./Hoffmeyer-Zlotnik, J./Krebs, D. (Hrsg.): *Gewichtung in der Umfragepraxis*. Westdeutscher Verlag, Opladen 1994, S. 112-123.
- Lüttinger, P./Riede, T.(1997): Der Mikrozensus: amtliche Daten für die Sozialforschung, *ZUMA-Nachrichten* Nr. 41, November 1997: 19-43.
- Meyer, Kurt (1994): Zum Auswahlplan des Mikrozensus ab 1990. In: Gabler, S./Hoffmeyer-Zlotnik, J./Krebs, D. (Hrsg.): *Gewichtung in der Umfragepraxis*. Westdeutscher Verlag, Opladen 1994, S. 106-111.
- Statistisches Bundesamt (Hrsg.) (2004): *Arbeitsunterlagen zum Mikrozensus. Das Erhebungsprogramm des Mikrozensus seit 1957*. Wiesbaden, November 2004.
- Statistisches Bundesamt (Hrsg.)(1989): *Inhaltliche Fragen bevölkerungsstatistischer Stichproben am Beispiel des Mikrozensus. Bericht zur Konferenz vom 21. und 22. Oktober 1988. Heft 10 der Schriftenreihe Ausgewählte Arbeitsunterlagen zur Bundesstatistik*. Wiesbaden 1989.
- Schmidt, S. (2000): *Erwerbstätigkeit im Mikrozensus. Konzept, Definition, Umsetzung*. ZUMA-Arbeitsbericht 2000/01.

Zur Weitergabe faktisch anonymisierter Einzeldaten:

- Müller, W./Blien, U./Knoche, P./Wirth, H. (1991): *Die Faktische Anonymität von Mikrodaten*, (Band 19 der Schriftenreihe Forum der Bundesstatistik). Metzler-Poeschel, Stuttgart 1991.
- Wirth, H. (1992): *Die faktische Anonymität von Mikrodaten: Ergebnisse und Konsequenzen eines Forschungsprojektes*. *ZUMA-Nachrichten* Nr. 30, Mai 1992.

Zur Fehlerrechnung im Mikrozensus:

- Statistisches Bundesamt 1973: *Bevölkerung und Kultur. Fachserie A, Reihe 6 Erwerbstätigkeit. I. Entwicklung der Erwerbstätigkeit 1973*. Stuttgart/Mainz: Kohlhammer.
- Krug, W./Nourney, M./Schmidt, J. (1994): *Wirtschafts- und Sozialstatistik. Gewinnung von Daten.*, München: Oldenbourg (3. völlig neubearb. Auflage)

Ergebnisse des Mikrozensus 1973:

- Statistisches Bundesamt 1973: *Bevölkerung und Kultur. Fachserie A, Reihe 6 Erwerbstätigkeit. I. Entwicklung der Erwerbstätigkeit 1973*. Stuttgart/Mainz: Kohlhammer.

Darüber hinaus sind Veröffentlichungen in der Zeitschrift 'Wirtschaft und Statistik' des Statistischen Bundesamtes zu finden.

Ein umfassendes Informationsangebot bietet das German Microdata Lab, ZUMA an:

<http://www.gesis.org/Dauerbeobachtung/GML/>

Publikationen, die auf den Scientific-Use-Files der Mikrozensen basieren finden sich auf der Webseite

http://www.gesis.org/Dauerbeobachtung/GML/Service/MZ-SUF-Literatur/MZ_Literatur.htm